DOCUMENT RESUME

ED 058 605 CG 007 024

AUTHOR Blanchfield, W. C.
TITLE College Dropout Identification, a Case Study.
INSTITUTION Utica Coll., N.Y.
PUB DATE [71]
NOTE 9p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Admission Criteria; *College Students; *Discriminant Analysis; Dropout Identification; *Financial Support; Potential Dropouts; Social Attitudes; Student Loan Programs
IDENTIFIERS MDA; Multiple Discriminant Analysis; SCS; Social Consciousness Score

ABSTRACT

This article presents a summary of some research on a statistical approach to identifying potentially successful and dropout students in a college population. The technique utilized is Multiple Discriminant Analysis (MDA). MDA eliminates the shortcomings of regression-correlation and uses only quantifiable information. The 2 variables that proved significant in the 3 tests were (1) the Social Consciousness Score test (administered by the Educational Testing Service) and (2) the percentage of college costs financed by grants. It was shown that successful students have a greater concern for social issues which is reflected in their greater persistance in college; also that successful students have a higher percentage of grants than unsuccessful students. High School Rank proved significant, while High School Average did not. Conclusions include that (1) MDA has application (69-8790 successful in study) for the college population as a whole; (2) loans are not necessarily the most effective way to finance higher education; and (3) the entire area of useful indicators of college student success should be reevaluated. (TA)

ED 058605

COLLEGE DROPOUT IDENTIFICATION, A CASE STUDY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

W. C. Blanchfield, Ph.D.
Assistant Professor of
Economics
Utica College

CG 007 024

INTRODUCTION

The process of selecting and evaluating High School students for the scarce openings in our Colleges goes on continuously. The process may be informal, interviews and personal judgment being the major ingredients; or it may be a highly structured mathematical process, most often using regression-correlation techniques. Neither method has been distinguished by its success.

The shortcomings of the subjective approach are obvious: too often the interview is hurried and the important indicators of possible success in college are not uncovered. The structured research has most often been an attempt to predict future quality point averages using prior data.

The reasons for the lack of previous success in the structured research are twofold: 1. regression-correlation is only suitable when the prior data is adequately separated; and 2. the premise upon which the technique is based is not suitable: that is, that it is possible to rank students in a significant manner to predict future success or failure.

Rather than ranking students on a scale, the emphasis should be on identifying potentially successful or dropout students. The technique ideally suited for this is Multiple Discriminant Analysis (MDA).

Multiple Discriminant Analysis (MDA) is a technique which eliminates the shortcomings of regression-correlation techniques. First, MDA places a dependent variable in a category rather than in a relative position and thus eliminates the problem of the independent variates being bunched too closely for regression techniques to rank them adequately. Secondly, since MDA uses only quantifiable information, personal interviewing techniques must be limited to factual data thus avoiding the gray area of subjective information.

MDA has proven successful in a variety of research efforts; the author used the technique to separate successful and unsuccessful students.

The research took place at Utica College, a branch of Syracuse University.

METHOD

A quantifiable student profile was the first objective of the study. A total of eight variables were chosen for the initial discriminant test and they are listed below.

TABLE I

Variables initially chosen

Variable	Source
Social Consciousness Score	College Student Questionnaire (given to all freshmen at Utica College)
Scholastic Aptitude Test - Verbal	Educational Testing Service (taken by students in High School)
Scholastic Aptitude Test - Math	Educational Testing Service (taken by students in High School)
College First Semester Grade Pt. Avg.	Records at Utica College
High School Average	" " " "
High School Rank (Quintile)	" " " "
Percentage of College Costs Financed by Loans	" " " "
Percentage of College Costs Financed by Grants	" " " "

The Social Consciousness Score ranges from 10 - 40. The Social Consciousness Score is administered by the Educational Testing Service. The test has a series of questions designed to indicate whether a student has an interest in matters of social concern. There is no attempt to uncover a student's political persuasions but rather whether he has an interest in such matters. High scores indicate that a student is aware of social problems (poverty, racial difficulties) and is concerned about them. Low scores indicate a lack of concern for such problems. In no way does it indicate political or social bias. The variable is particularly suited for MDA because it is the only quantifiable score relating a student's interest in the world around him.

The Scholastic Aptitude Tests (Verbal and Mathematics) are two of the primary indicators that admission counselors use to decide on the acceptance of students. They were inserted in the model to test whether they should warrant such importance.

Previous studies in predicting college dropouts indicate that first semester grade point average is the most significant indicator in a student's profile.[1] The model will test this premise.

The High School Rank and High School Average are also common criteria used to admit students into college. How effective they are in identifying future dropouts has never been studied.

The percentage of college costs financed by loans and the percentage of costs financed by grants are the two economic variables in the study. They are there to find the best method of financing an individual's education. Most experts in the field have advocated loans as the cheapest and most effective way of keeping students in college.[2] The model will test this assumption.

Since all the variables are quantifiable they are acceptable in an MDA program.[3] The mathematics of the technique is listed below.

Mathematics of MDA

Assume that there is a set of scores for each group; which linear combination of scores will best discriminate between the groups?

Assume that there are N normally distributed observations on p variables X_i which are denoted by X_{it} (i = 1, 2 . . . p; t =1, 2 . . . N).

---

1 For a comprehensive view of studies on student characteristics, see John R. Hills, "Admissions Procedures that Make Sense", The American College, (New York: Wiley and Sons, 1962), pp. 16 - 24.

2 Both Seymour Harris and Otto Eckstein have advocated loans as a means of retaining students in college. See Seymour Harris, editor, Higher Education in the United States, (Cambridge: Harvard University Press, 1960), p. 9.

3 The mathematics was developed by R. A. Fisher in "The Use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics, no. 7.

Classify these into two groups for t = 1, 2 . . . N_1 and t = N_1 + 1, N_1 + 2, . . . N_1 + N_2 = N. The means in each group are defined by:

(1) $\bar{X}_{*i} = \sum_{t=1}^{N_1} \frac{X_{it}}{N_1}$, $\bar{X}_i^{**} = \sum_{t=N_1+1}^{N} \frac{X_{it}}{N_2}$ (i = 1, 2 . . . p)

Let the difference of the means be:

(2) $d_i = \bar{X}_i^{**} - \bar{X}_i^{*}$ (i = 1, 2 . . . p)

The object is to find the linear function of the difference of the means:

(3) $Z = k_1 d_1 + k_2 d_2 + \ldots k_p d_p$

which best discriminates between the two sets of variables. Its square should be a maximum relative to its variance. The variance is proportional to:

(4) $Q = \sum_{i=1}^{p} \sum_{j=1}^{p} k_i k_j S_{ij}$

where k = coefficient of each variable and S = each variable in the model.

In order to maximize Z^2 (3) under the condition that the variance (4) is a constant, the following function is formed:

(5) $F = Z^2 - \lambda Q = \sum_{i=1}^{p} \sum_{j=1}^{p} k_i k_j d_i d_j - \lambda\, k_i k_j S_{ij}$

where λ is a Lagrange multiplier.

Expression (5) is differentiated partially with respect to k_j (j = 1, 2, . . . p) and the following set of equations is obtained:

(6) $d_j \sum_{i=1}^{p} k_i d_i = \lambda \sum_{i=1}^{p} k_i S_{ij}$ (J = 1, 2, . . . p)

To simplify let $\lambda = \sum_{i=1}^{p} k_i d_i$ and the following is obtained:

$S_{11}k_1 + S_{12}k_2 + \ldots + S_{1p}k_p = d_1$

$S_{12}k_1 + S_{22}k_2 + \ldots + S_{2p}k_p = d_2$

. . . . . .

$S_{1p}k_1 + S_{1p}k_2 + \ldots + S_{pp}k_p = d_p$

The solutions k_i are proportional to the estimates of the coefficients of the linear function which in the population corresponding to the sample

discriminates best between the groups.

A test of significance is also included in the solutions. The test uses the F distribution and the generalized student distribution in the following manner:

$$R^2 = \frac{N_1 N_2 (k_1 d_1 \ \ldots \ k_p d_p)}{N}$$

then the variance ratio:

$$F = \frac{(N - p - 1)\ R^2}{p(1 - R^2)}$$

is the F distribution for $n_1 = p$ and $n_2 = N - p - 1$ degrees of freedom. In this way the hypothesis may be tested that the discriminant function was a chance occurrence.

Results of Identification Model

In all three tests the students from both categories (dropouts and successful students) were chosen randomly from the files of Utica College. After the relevant information was obtained for each student, the entire group was analyzed by the MDA model. The model then identified successful and dropout students. The results of such identification are listed below.

Table II

Variables and F Test Results for Three Discriminant Tests

Variables	Test 1	Test 2	Test 3
1. Social Consciousness Score	8.101*	15.200*	16.321*
2. Percentage of College Costs Financed by Grants	16.924*	14.240*	13.124*
3. First semester grade point average	20.967*	36.842*	51.768*
4. High School Rank (Quintile)	8.885*	17.166*	25.839*
5. Scholastic Aptitude Test Verbal Score	1.358		
6. Scholastic Aptitude Test Mathematics Score	0.671		
7. Percentage of College Costs Financed by Loans	2.259	0.488	
8. High School Average	1.799		

Table II - continued

Variables	Test 1	Test 2	Test 3
Number and Percentage of Active Students correctly identified	25(69.4%)	32(80.0%)	33(86.0%)
Number and Percentage of Inactive Students correctly identified	29(78.3%)	35(87.5%)	32(82.0%)

*Significant at .05 level

The students were drawn randomly, a different group for each test, from the files of Utica College.

The Social Consciousness Score proved significant in all three tests. This means that one can identify by this variable a dropout or successful student and that there is significant difference in social awareness between students who stay in college and those who drop out. The successful students have greater concern for social issues and this is reflected in their greater persistence in college.

The percentage of college costs financed by grants was also significant in all three tests. Thus, successful students have higher percentage of grants than unsuccessful students. One is tempted to assume that the result follows from the fact that grants are given to more successful students. This was not proven, however, by subsequent correlation tests run on all eight variables.[4]

A more likely explanation is that the awarding of a grant provides a degree of security to a student, thus providing more incentive to remain in college.

The fact that the percentage of college costs financed by loans was not significant disputes the hypothesis set down by Harris and Eckstein.[5]

---

[4] The correlation coefficient between the High School Average and % of Grants was .261. Between the First Semester Grade Point Average and the Percentage of Grants was only .291.

[5] OP. Cit; p. 9.

The awarding of a loan is often looked upon unfavorably by the recipient. The worry incurred over the rising debt is also a problem for such students.

Of the two high school indicators, the High School Rank proved significant while the High School Average was not significant. The rank is biased in that a good student in stiff competition may have a low rank while a poorer student may rank high in a less competitive situation. The High School average should account for this, however, but this proved not to be significant in the testing.

The first semester college grade point average was, expectedly, significant in all three tests. There is nothing like a good academic beginning to give a student confidence.

The two aptitude tests, Verbal and Mathematics, were surprisingly not significant. This could be due to the lack of separation of the scores (most students at Utica College range from 430 - 480 in scores) or to the fact that at the marginal schools such as Utica, the score is not as useful as at other more prestigious schools. Whichever the reason, the aptitude tests were not a factor in identifying college dropouts.

Although one should not generalize for the entire college population, the results of the study at Utica College show clearly that the faith of admissions counselors in High School averages and the Aptitude Tests is open to question. Likewise, the economic assumption discussed earlier, that loans are the most efficient method of financing higher education expenses, is not verified by the study at Utica College.

The MDA model has clearly shown that grants are identified with successful students. Loans, moreover, are not necessarily associated with dropouts, but are not significantly associated with either group.

Conclusions

Three significant conclusions result from the research on dropouts at Utica College. One is that the statistical technique, Multiple Discriminant Analysis,

proved highly successful (69 - 87%) in identifying dropouts at the college. This is remarkable in view of the rather poor results that have been recorded in previous correlation tests.[6]

While the analysis was only applied to the Utica College population there is application for the college population as a whole. The method is certainly applicable -- the variables may be chosen to suit the particular college. The important point is to uncover the relevant variables for the particular college. MDA is a perfect tool for such analysis.

Secondly, a number of economic policy conclusions emerge from the results, the most important of which is that loans are not necessarily the most effective way to finance higher education.

Thirdly, the entire area of useful indicators of student success in college should be reevaluated. There is some question as to whether High School Grades and Achievements Tests as criteria deserve the attention they get from Admissions Counselors.

Future research should be directed toward clarifying the issues raised here. Multiple Discriminant Analysis is an ideal method for such research.

---

[6]Studies conducted by the Dean's Office at Utica College yielded correlation coefficients around .08. This means that only .06% of the variance was explained by the independent variable.